Crystal structure of the 25 kDa subunit of human cleavage factor I$_m$

Molly Coseno[1], Georges Martin[2], Christopher Berger[3], Gregory Gilmartin[1], Walter Keller[2] and Sylvie Doublié[1,*]

[1]Department of Microbiology and Department of Molecular Genetics, University of Vermont, Burlington, VT 05405, USA, [2]Department of Cell Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland and [3]Molecular Physiology and Biophysics, University of Vermont, Burlington, VT 05405, USA

ABSTRACT

Cleavage factor I$_m$ is an essential component of the pre-messenger RNA 3′-end processing machinery in higher eukaryotes, participating in both the polyadenylation and cleavage steps. Cleavage factor I$_m$ is an oligomer composed of a small 25 kDa subunit (CF I$_m$25) and a variable larger subunit of either 59, 68 or 72 kDa. The small subunit also interacts with RNA, poly(A) polymerase, and the nuclear poly(A)-binding protein. These protein–protein interactions are thought to be facilitated by the Nudix domain of CF I$_m$25, a hydrolase motif with a characteristic α/β/α fold and a conserved catalytic sequence or Nudix box. We present here the crystal structures of human CF I$_m$25 in its free and diadenosine tetraphosphate (Ap$_4$A) bound forms at 1.85 and 1.80 Å, respectively. CF I$_m$25 crystallizes as a dimer and presents the classical Nudix fold. Results from crystallographic and biochemical experiments suggest that CF I$_m$25 makes use of its Nudix fold to bind but not hydrolyze ATP and Ap$_4$A. The complex and apo protein structures provide insight into the active oligomeric state of CF I$_m$ and suggest a possible role of nucleotide binding in either the polyadenylation and/or cleavage steps of pre-messenger RNA 3′-end processing.

INTRODUCTION

Pre-messenger RNA 3′-end processing in eukaryotes is a two-step reaction consisting of endonucleolytic cleavage of the pre-mRNA followed by addition of a poly(A) tail at the 3′ end of the upstream cleavage product (1–3). The coupling of these processing reactions relies on multiple protein–protein and protein–RNA interactions. The factors that are necessary and sufficient to reconstitute cleavage and polyadenylation in a mammalian *in vitro* system are poly(A) polymerase (PAP), cleavage and polyadenylation specificity factor (CPSF), cleavage stimulation factor (CstF), cleavage factor I$_m$ (CF I$_m$), cleavage factor II$_m$ (CF II$_m$) and the nuclear poly(A)-binding protein 1 (PABPN1). The recruitment of the mammalian polyadenylation machinery to the pre-mRNA relies on the recognition of conserved sequence elements, such as the highly conserved hexamer AAUAAA recognized by CPSF and a U-rich sequence downstream of the cleavage site recognized by CstF. Additionally a set of UGUA elements, a third sequence element found upstream of the cleavage site and not as universally conserved, is recognized by CF I$_m$ (4,5). Upon completion of the processing steps the mRNA can be efficiently transported out of the nucleus into the cytoplasm. Homologs of CF I subunits are also found in cDNA databases of *Caenorabditis elegans*, *Drosophila melanogaster* and plants.

CF I$_m$ is an oligomer composed of a small 25 kDa subunit (CF I$_m$25, also referred to as CPSF5 or NUDT21) and a larger subunit of either 59, 68 or 72 kDa (6). The three larger subunits share substantial sequence homology. They are encoded on two different genes and the 72 kDa subunit is a splice variant of the 68 kDa polypeptide. CF I$_m$ can be reconstituted *in vitro* from the 25- and 68 kDa subunits (6). CF I$_m$, in addition to having a role in regulation of poly(A) site recognition, is also involved in stabilizing CPSF at the conserved hexamer, enhances the rate of poly (A) site cleavage *in vitro*, and has been shown to interact with PAP and PABPN1 via its 25 kDa subunit (7). Furthermore CF I$_m$ also interacts with splicing factors, indicating a role in communicating between different RNA-processing complexes (8).

Much of what we know about complex formation of processing factors on the pre-mRNA and its regulation

*To whom correspondence should be addressed. Tel: +1 802 656 9531; Fax: +1 802 656 8749; Email: sdoublie@uvm.edu

Figure 1. Sequence alignment of CF I_m25 with Nudix proteins. ClustalW sequence alignment of CF I_m25 and Nudix proteins of known structure (44). PDB ID codes are shown to the right of the enzyme names followed by the residue number of the first amino acid. The position of the Nudix box is indicated below the alignment as a grey/black bar where the black part marks the position of helix α2. Residues are on a light blue background if over 70% conserved and are on yellow background if invariant. The two catalytic glutamates conserved in most Nudix enzymes are displayed in red font on light orange background. L124 and I128, which are found in place of the conserved glutamates, are boxed in the CF I_m25 sequence. The abbreviations are defined as: PPHase for pyrophosphohydrolase and PPase for pyrophosphatase.

has been investigated biochemically. Structural characterization of individual and multicomponent processing factors will elucidate the domain interactions important for the 3′ pre-mRNA processing mechanism that cannot be deciphered by biochemical means. The structure of the 25 kDa subunit of CF I_m (CF I_m25) will help further define the domains important for substrate and protein interactions.

The structure of CF I_m25 will also allow us to investigate the function of a Nudix domain present in this protein. The first Nudix protein to be characterized enzymatically and structurally was *Escherichia coli* MutT (9). Nudix proteins are generally characterized as housekeeping enzymes due to their role in hydrolysis of substrates described as **nu**cleoside **di**phosphate linked to another moiety **X**, many of which are potentially toxic molecules (10,11). Nudix proteins have a conserved Nudix fold consisting of an α/β/α sandwich. Within the Nudix fold the consensus sequence of the Nudix box is $GX_5EX_7REUXEEXGU$, where U is a hydrophobic residue and X is any residue, and folds into a loop–α helix–loop structure (9). Interestingly, the Nudix box in CF I_m25 lacks two of the four conserved glutamate residues, three of which were shown to be important for catalysis (Figure 1) (12).

In this study we present the crystal structure of human CF I_m25 alone and in complex with diadenosine tetraphosphate (Ap_4A). CF I_m25 crystallizes as a dimer, which is also the oligomeric state of the protein in solution. The CF I_m25 structure presents the classic Nudix α/β/α fold and harbors residues outside of the Nudix core that could potentiate ligand binding. Structural and biochemical evidence suggests that CF I_m25 binds, but does not hydrolyze, mono and di-adenosine nucleotides.

MATERIALS AND METHODS

Protein purification

The construction of the plasmid expressing the 25 kDa subunit of human cleavage factor I_m (CF I_m) with an N-terminal 6xHis tag was previously described (7). The cDNA of CF I_m25 was subsequently cloned into a Gateway® vector with a dual 6xHis-maltose binding protein (MBP) affinity tag provided by Dr David S. Waugh (National Cancer Institute, Frederick, MD) (13). The following primers were used in the polymerase chain reaction (PCR): 5′-GAG AAC CTG TAC TTC CAG GGT ATG TCT GTG GTA CCG CCC-3′, 5′-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTT ATT AGT TGT AAA TAA AAT TGA A-3′ and 5′-GGGG ACA AGT TTG TAC AAA AAA GCA GGC TCG GAG AAC CTG TAC TTC CAG-3′. CF I_m25 (227 amino acids) was then expressed in Rosetta (DE3) pLysS cells (Novagen) and grown in LB medium for 24 h at 25°C following induction with 0.4 mM IPTG. The protein was purified on Nickel NTA beads (Qiagen), followed by a Tobacco etch virus (Tev) protease cleavage step, and cation exchange on a Resource S column (GE Healthcare): Cells were lysed at 4°C by sonication in a buffer containing 20 mM Tris–HCl pH 8.0, 200 mM NaCl, and a protease inhibitor tablet (Roche). The lysate was centrifuged at 12 000 × g and incubated with Ni-NTA beads (Qiagen) for 1 h at 4°C. Protein was eluted with lysis buffer containing 100–500 mM imidazole. Pooled fractions were dialyzed into 10% (v/v) glycerol, 20 mM Tris–HCl pH 7.5 and 50 mM KCl. CF I_m25 was cleaved from the MBP tag by the addition of equimolar amounts of Tev protease. MBP and Tev were then separated from CF I_m25 by elution with a 50 mM–1 M KCl gradient at pH 7.5 on a Resource S column (GE Healthcare). The selenomethionyl protein was produced by inhibiting methionine biosynthesis in *E. coli* (14) and purified as described earlier. Proteins were concentrated to about 6–20 mg/ml (Millipore Amicon Ultra-15), flash frozen and stored at −80°C.

Crystallization

CF I_m25 crystals were initially obtained with the sitting drop method in a 96-well tray format. Sitting drops were set up with a 925 PC Workstation (Gilson) by mixing

Table 1. Data collection and refinement statistics

Data collection	Se-Peak	Se-Edge	Se-Remote	Ap$_4$A complex
Wavelength (Å)	0.97923	0.97939	0.97166	0.9
Resolution (Å)	30–1.82 (1.89–1.82)[a]	30–1.85 (1.92–1.85)[a]	30–1.90 (1.97–1.90)[a]	15–1.80 (1.86–1.80)[a]
Number of reflections				
Measured	157128	148933	136117	543233
Unique	23262	22468	20608	25257
Redundancy	6.8 (5.6)	6.6 (5.2)	6.6 (5.2)	21.6 (6.7)
Completeness (%)	97.7 (85.9)	96.5 (80.6)	95.7(79.9)	99.8 (98.2)
R_{merge} (%)[b]	8.6 (51.7)	7.3 (57.7)	7.0 (52.8)	13.8 (26.0)
I/σ	16.09 (2.2)	19.7 (2.3)	19.3 (2.4)	66.0 (4.4)
Selenium sites	3			
R_{cullis}[c]	0.788	0.886	0.908	
Phasing power	1.13	0.736	0.815	
Overall mean FOM	0.33/0.72[d]			
PDB ID Code	3BAP			3BHO
Number of atoms				
Protein (non-hydrogen)	1649			1663
Water	167			169
R_{work}[e]	0.212			0.202
R_{free}[e]	0.234			0.225
r.m.s.d. bond distances (Å)	0.005			0.004
r.m.s.d. bond angles (°)	1.320			1.269
Ramachandran plot				
Most favored region (%)	92.0			92.9
Additionally allowed (%)	8.0			7.1
Generously allowed (%)	0			0
Disallowed (%)	0			0
Average B factor all atoms (Å^2)				
Wilson B-factor	33.3			31.4
Protein only	38.7			33.0
Water only	50.2			44.7
Ligand only	58.7			50.4

[a]High resolution shell is shown in parentheses.
[b]$R_{merge} = \sum|I - <I>|/\sum I$, where $<I>$ is the average intensity from multiple observations of symmetry-related reflections.
[c]$R_{Cullis} = [<(LOC)^2>]^{1/2} (<|\Delta F|^2>)^{1/2}$, where LOC is the lack-of-closure error.
[d]Before and after density modification.
[e]R_{work} and $R_{free} = \sum_h ||F_o| - |F_c||/\sum_h |F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes. R_{free} was calculated with 10% of the reflections not used in refinement.

0.6 µl of protein (16 mg/ml) with 0.6 µl reservoir solution (25% w/v PEG 3350, 0.2 M MgCl$_2$, 0.1 M Tris–HCl pH 8.5) (Hampton Research Index Screen, condition 85) and equilibrated against 160 µl reservoir buffer. Subsequently, crystals were obtained by streak seeding hanging drops with a protein concentration of 6 mg/ml. Hanging drops were set up by mixing 1 µl of protein with 1 µl of reservoir solution under the same crystallization conditions. Trigonal crystals grew to a maximum length of 200 µm in space group P3$_1$21 (P3$_2$21) with unit cell parameters $a = b = 80.11$ Å, $c = 72.21$ Å and $\gamma = 120°$. There is one molecule per asymmetric unit with an estimated solvent content of 52%. Crystals of the complex with Ap$_4$A were obtained with the hanging drop method. The hanging drops were set up manually by mixing 1 µl of protein with 1 µl of reservoir solution (25% w/v PEG 3350, 0.025 M MgCl$_2$, 0.1 M Tris–HCl pH 7.5). The drops were streak seeded after a 24 h incubation period. When the crystals reached at least 100 µm, Ap$_4$A at a final concentration of 44 mM was added directly to the drop. After 6 h the soaked Ap$_4$A crystals were cryoprotected by the addition of 1 µl of 25% (w/v) PEG 3350 and 50% (v/v) glycerol to the 2 µl hanging drop prior to flash cooling in liquid nitrogen.

Crystallographic data collection

Multiple wavelength anomalous diffraction (MAD) data were collected at beamline 23-ID-D (Advanced Photon Source at Argonne National Laboratory) on a MAR m300 CCD detector. One complete selenomethionyl MAD dataset was collected on one crystal at the peak, inflection and high-energy remote wavelengths to a maximum resolution of 1.85 Å. Data were collected at 1.80 Å resolution on the Ap$_4$A complex at beamline X12B (National Synchrotron Light Source) on a Quantum-4 CCD (ADSC) detector. The data from three Ap$_4$A-soaked crystals were merged to increase redundancy. Diffraction data were processed and scaled with DENZO and SCALEPACK (15). Data collection statistics are summarized in Table 1.

Structure determination and refinement

The program SOLVE (16) identified three of the four selenium sites (the N-terminal methionine is disordered or

missing). AutoSHARP (17) was then used for refinement of the selenium parameters. The space group was judged to be P3$_1$21 and not the enantiomorphic P3$_2$21, based on the map quality and continuity. The phasing information was then used in RESOLVE (18) for density modification and iterative model building. Seventy percent of the model was built by RESOLVE. The remaining residues were built manually using the program COOT (19). Residues 1–20 and 132–135 were omitted from the model because of poorly defined density.

Iterative rounds of refinement including simulated annealing, energy minimization, and B-factor refinement were done with CNS (20). Each refinement round was followed by rebuilding in COOT. A composite simulated annealing omit map was generated in CNS to validate the model and build the remaining side chains. Water molecules were added with CNS and COOT. The quality of the model was evaluated with PROCHECK (21). All non-glycine residues fall within either the most favored or additionally allowed regions of the Ramachandran plot. The refined model of the unliganded protein provided phases to calculate an isomorphous difference Fourier (Fo–Fo) map between the unliganded protein and the complex with Ap$_4$A (22). The R_{cross} on amplitudes between the two datasets is 0.149, indicating good isomorphism between the two crystals. The resulting map showed clear density for one adenine base and three phosphates of Ap$_4$A. The complex with Ap$_4$A was refined with CNS (20). The refinement statistics for both structures are reported in Table 1. Figures were drawn with PyMOL (23).

Oligomeric state determination

Size exclusion chromatography was performed with a Superdex 75 column (GE Healthcare). The protein sample or molecular mass standards were applied to the Superdex 75 column and eluted with 10% glycerol, 20 mM Tris pH 7.5 and 50 mM KCl. Standard proteins (Sigma) were: lysozyme (MW 14.4 kDa), egg albumin (45 kDa), bovine serum albumin (67 kDa) and the void volume was determined with blue dextran (GE Healthcare). Dynamic light scattering was also used to confirm the oligomeric state of CF I$_m$25 (Dynapro, Wyatt). The buried surface area (2800 Å^2) was calculated with AREAIMOL (24) and with CNS (20) (2700 Å^2).

Hydrolase activity assay

The standard reaction mixture contained 50 mM Tris–HCl, pH 7.5 or 8.5, 50 mM KCl, 5 mM MgCl$_2$, 1 mM DTT, 20 μM CF I$_m$25, 2 U calf intestinal phosphatase (New England Biolabs) or 1 U *S. cerevisiae* pyrophosphatase (Sigma), and 2 or 4 mM substrate in a total volume of 50 μl. The putative substrates tested for CF I$_m$25 hydrolase activity include ATP, Ap$_4$A (Sigma) and m^{7}G (5′)ppp(5′)G cap structure analog (New England Biolabs). The reactants were combined at 4°C then incubated at 25°C for 30 and 60 min. The reaction was terminated by the addition of 250 μl of 20 mM EDTA and the liberated orthophosphate was determined by the colorimetric assay of Ames and Dubin (9,25). The limit of detection of this assay is 5 μM of orthophosphate.

Fluorescence measurements

Steady-state tryptophan fluorescence was measured with a Quantamaster fluorimeter (Photon Technology International, South Brunswick, NJ) as described (26) with a WG320 cut-off emission filter. CF I$_m$25 contains four tryptophans and only three are built in the structure: Trp148 and Trp149 are within 10–15 Å of the active site and Trp139 is located within 20 Å of the active site. The fourth tryptophan, Trp13, is located in the disordered portion of the amino terminus. The tryptophan emission spectrum was measured by excitation of the samples at 295 nm and collecting the emitted fluorescence at 90° to the incident light over the range 300–400 nm. The slit widths were set at a resolution of 1 nm for excitation and 4 nm for emission. Fluorescence measurements of all protein samples were performed using a microcuvettte with a magnetic stir bar in 20 mM Tris–HCl pH 7.5, 50 mM KCl and 25 mM MgCl$_2$ at 25°C for the protein alone and in the presence of increasing amounts of nucleotide. All fluorescence measurements were corrected for Raman scatter and background fluorescence and represent experiments performed in triplicate and then normalized and averaged. The ATP fluorescence data were fit with a single hyperbola [$y = ax/(b + x)$] with a K_d of 1.53 ± 0.18 mM (1.17–1.89 mM at 95% confidence interval). The Ap$_4$A data were fitted with a single hyperbola [$y = ax/(b + x)$] with a K_d of 2.44 ± 0.49 μM (1.46–3.43 μM at 95% confidence interval).

RESULTS

Crystal structure of human CF I$_m$25

The original N-terminal His-tagged plasmid of human CF I$_m$25 (7) did not express to high enough levels for structural studies. We therefore inserted the coding sequence of CF I$_m$25 into a dual HisMBP vector (13). With the resulting expression vector, 1.5 mg of protein could be purified from 1l of culture. The structure was solved to a resolution of 1.85 Å by multiple wavelength anomalous diffraction of the selenomethionyl protein variant. Residues 21–131 and 136–227 are visible in the electron density map. A complex with Ap$_4$A was also obtained and refined to a resolution of 1.80 Å.

Description of the structure

Human CF I$_m$25 is composed of 227 residues, with a calculated molecular weight of 26 kDa. CF I$_m$25 elutes as a dimer in gel exclusion chromatography with an apparent molecular weight of ~53 kDa (Supplementary Figure). The dimeric state of CF I$_m$25 has also been confirmed by dynamic light scattering. In the crystal structure, dimer formation relates two monomers by a 2-fold crystallographic axis. The Nudix domain is located in the middle of the protein and spans residues 77–202 (Figure 2). The Nudix fold comprises two mixed β sheets flanked by two helices (α2 and α3). The Nudix box is located in helix α2
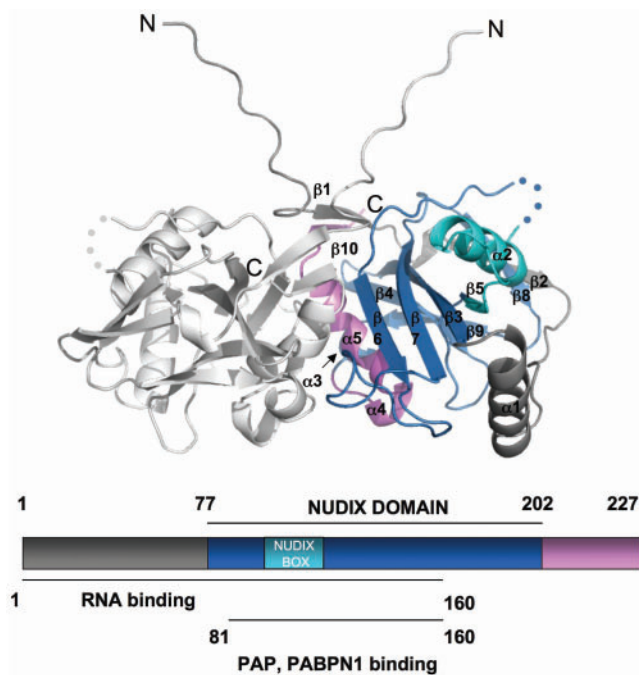
Figure 2. Domain organization of CF I$_m$25. Ribbon diagram of the CF I$_m$25 dimer comprising residues 21–131 and 136–227. The ribbon color scheme corresponds to that of the domain architecture shown below. The second monomer on the left is shown in light grey. The secondary structure numbering is based on a DSSP analysis (45).

(residues 109–131) and the preceding loop. The region immediately preceding the Nudix domain (residues 21–76) folds into two distinct segments: residues 21–32 form an extended loop structure that projects away from the globular domain of the protein (residues 1–20 are not visible in the electron density map). The extended loop is stabilized via crystal-packing interactions with a symmetry-related monomer and this region, which has been shown to participate in RNA binding, is likely to adopt a different fold *in vivo* (7). Lys23 has previously been shown to undergo acetylation, a modification which modulates the interaction of CF I$_m$25 with PAP (27). The second segment of the N-terminal region (residues 33–76) contributes two short β strands and a long helix. Within this region an additional modification, this time phosphorylation, was reported for Tyr40 (28). We do not see either modification because the protein was expressed in *E. coli*. The C-terminal region (residues 203–227) is composed of two helices (α4 and α5) and a short beta strand (β10) comprising the final four residues.

Structural homologs of CF I$_m$25 were searched with the DALI server (29) and significant structural homology was found, but only with other Nudix proteins, including *Deinococcus radiodurans* DR1025 [*Z* score = 11.2, PDB ID code 1SJY; (12)], Ap$_4$A hydrolase [*Z* score = 10.0, PDB ID code 1KT9; (30)], ADP ribose pyrophosphatase [*Z* score = 9.3; PDB ID code 1V8M; (31)] and the mRNA decapping enzyme DCP2P [*Z* score = 9.0; PDB ID code 2A6T; (32)]. The highest *Z*-score (33.6 with associated rmsd of 0.6 Å) was actually that of another crystal structure of CF I$_m$25 for which coordinates were deposited

but had no publication associated with it (Structural Genomics Consortium, Karolinska Institute; PDB ID code 2CL3).

Dimer interactions

Dimer formation buries an interface of about 2700 Å^2 (20,24). The buried surface is extensive and represents 13% of the surface of the homodimer. The dimer interface is maintained by approximately 20 residues per monomer, which participate in both hydrogen bonding and non-polar contacts. The monomer interactions are facilitated by the amino terminal extension (residues 21–32) and by helix α5, the loop preceding helix α5, helix α6, beta strand β6, the loop linking β6 and β7 and beta strand β10.

The main hydrogen-bonding contributions are between the amino terminal extension, beta strand β6 and beta strand β10. These involve residues Thr32 of chain A and Asp142 of chain B, Ser220 of chain A and Asn147 of chain B as well as the reverse monomer interactions. The loop preceding helix α5 and the loop between β6 and β7 contribute to dimer formation through hydrophobic contacts involving residues Pro159 and Tyr160 from chain A and Tyr202 and Phe199 from chain B. Additional interactions are provided by stacking helix α5 from one monomer on to strand β6 of the opposing monomer in a tail-to-tail fashion allowing hydrogen bonding between residues in the loop following helix α5. These hydrogen-bonding interactions include Arg221 and its symmetry mate and the carboxyl group of Ser220 with Asn147 from the opposing monomers.

Metal binding

CF I$_m$25 is missing the second and fourth conserved glutamates of the Nudix box (residues 124 and 128), which are replaced by a leucine and isoleucine, respectively. These glutamate residues are very often involved in metal binding (12). We therefore set out to investigate whether CF I$_m$25 was still capable of binding metal via the two remaining carboxylates. Although the CF I$_m$25 crystals were obtained in the presence of 200 mM MgCl$_2$, there was no identifiable Mg^{2+} bound in the electron density map. Magnesium is a light atom, which is usually not easily identifiable in electron density maps and can often be mistaken for a water molecule. We therefore used metals that are more electron dense such as MnCl$_2$ and GdCl$_3$ to identify putative metal site(s) (33). Mn^{2+} and Gd^{3+} present the added advantage that they are anomalous scattering atoms, which should allow unambiguous identification of metal sites. We were unable to identify binding of either metal, regardless of whether the metal was co-crystallized or soaked into pre-existing crystals. Since Nudix enzymes require a divalent cation for catalysis (9), these experiments suggested that it is unlikely that CF I$_m$25 functions as a hydrolase. This finding called for further investigation of CF I$_m$25's binding to potential substrates (see below).

Enzymatic assays and substrate binding

We next set out to investigate whether CF I$_m$25 binds a substrate. We used three different methods: a colorimetric

assay to measure the release of inorganic phosphate (25), co-crystallization/crystal soaking with putative substrates and steady-state tryptophan fluorescence experiments.

A colorimetric assay to measure the release of inorganic phosphate, and thus the potential hydrolytic activity of CF I_m25 on nucleotides, was performed (25). The following putative substrates were tested: ATP, Ap_4A and the $7mG(5')ppp(5')G$ cap analog. Ap_4A was chosen because the two enzymes found to be most structurally similar to CF I_m25 bind Ap_4A (12,30). ATP and the cap analog were tested because of their prominent role in RNA processing. The assay indicated that, within the limits of detection of the assay, none of the putative substrates tested were hydrolyzed.

Although CF I_m25 does not seem to possess a hydrolytic activity, it could still bind nucleotides. Several putative ligands were either co-crystallized or soaked into the crystals: ATP, GTP, ADP, GDP, diadenosine triphosphate (Ap_3A), Ap_4A, AMP, $7mG(5')ppp(5')G$ cap analog, NAD+ and GDP-mannose. In addition to testing these nucleotides, which are known substrates for Nudix enzymes, we also attempted to co-crystallize CF I_m25 with a 21mer RNA derived from the PAP γ cDNA sequence (4,5). Of all the putative ligands tested, only Ap_4A bound to the crystal (see later). The original crystallization conditions contained 200 mM $MgCl_2$ and 25% (w/v) polyethylene glycol (PEG) 3350. Because nucleotides can precipitate in the presence of high concentrations of PEG and magnesium (34), care was taken to modify the crystallization conditions so as to decrease or completely eliminate $MgCl_2$ and therefore lessen the risk of the nucleotide precipitating out of solution. Conditions with NaCl or even no salt could be used in lieu of $MgCl_2$ to grow unliganded crystals. Even when the divalent cation was omitted from the crystallization solution crystals did not form with any of the ligands tested (with the notable exception of Ap_4A), demonstrating that the lack of binding was not due to the nucleotide falling out of solution.

Binding studies were then performed employing the intrinsic tryptophan fluorescence properties of CF I_m25. The following Nudix ligands were tested: ATP, GTP, Ap_4A, inositol hexaphosphate (IP6), $7mG(5')ppp(5')G$ cap analog, ADP and AMP. Dissociation constants (K_d) measured by titrating the ligand concentration were estimated for ATP and Ap_4A (Figure 3A and B, respectively). The ATP data were fit with a single hyperbola curve, indicating one binding affinity with a K_d value of 1.53 ± 0.18 mM. The Ap_4A-binding data were fit to a single hyperbola curve and represent one binding site per monomer, resulting in a K_d of 2.44 ± 0.49 μM.

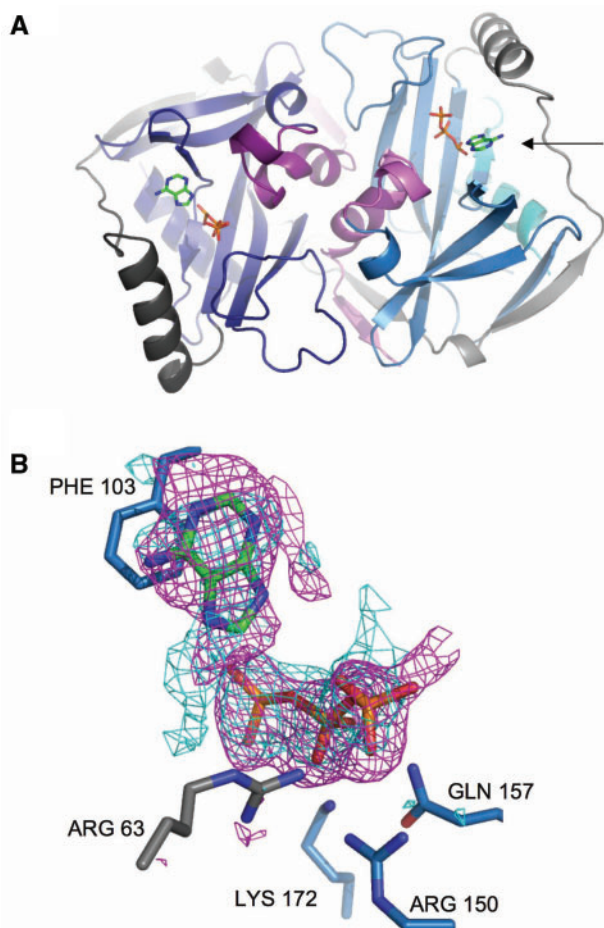## Complex of CF I_m25 with Ap_4A

Co-crystallization experiments were performed in order to resolve the ligand-binding interactions of CF I_m25. Of all the molecules screened only Ap_4A was captured in a co-crystal (Figure 4A). The Ap_4A co-crystal structure was determined by using the phases from the unliganded structure. The resulting unbiased isomorphous difference Fourier (Fo–Fo) map revealed distinct density for one of
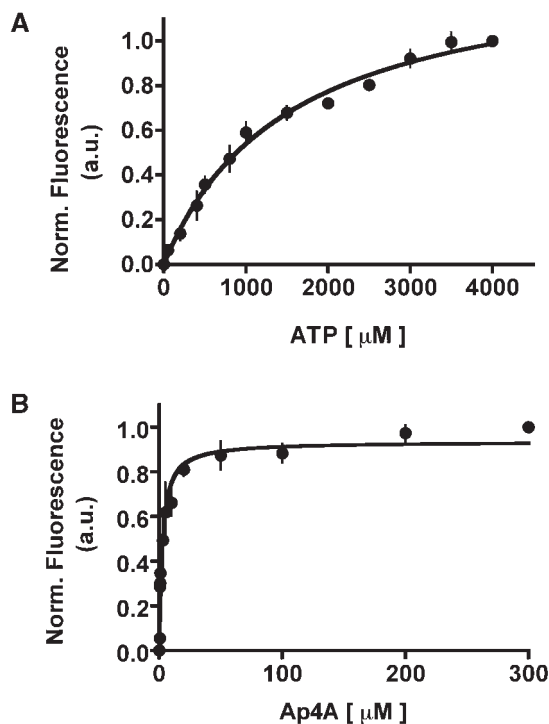
Figure 3. CF I_m25 steady-state tryptophan fluorescence experiments with ATP and Ap_4A. All data are represented as normalized and averaged experiments done in triplicate. Error bars that are not represented lie within the symbol. (**A**) ATP steady-state tryptophan fluorescence data are fit with a single hyperbola, with $K_d = 1.53$ mM. (**B**) Ap_4A steady-state tryptophan fluorescence data are fit with a single hyperbola, with $K_d = 2.44$ μM.

the adenine bases and three of the four phosphates of Ap_4A, in the cavity of CF I_m25 (Figure 4B). The rmsd between the bound and apo structures is low (0.41 Å), indicating that only very small changes take place upon substrate binding (20). The binding site residues Arg63, Arg150, Gln157 and Lys172 are found outside of the Nudix box region, within 3 Å of the triphosphate moiety, and are involved in coordinating the triphosphate moiety (Figure 4B). The binding site is composed of residues from helix α1, beta strand β6, the loop linking β6 and β7 and beta strand β7. In CF I_m25, Arg63 and Arg150 are highly conserved across species while Gln157 and Lys172 are moderately conserved. The majority of hydrogen-bonding interactions with Ap_4A involve the oxygens of the β- and γ phosphates. The ordered adenine base of the Ap_4A molecule stacks with Phe103, a residue contained within the Nudix domain. We note that the position of the γ phosphate of Ap_4A coincides with the position of a sulfate ion reported for CF I_m25 (Structural Genomics Consortium, Karolinska Institute; PDB ID code 2J8Q).

CF I_m25 has a core structure similar to that of the *D. radiodurans* Nudix protein DR1025 (12) (PDB ID code 1SU2) (Figure 5). Variations between these two structures arise from an extension in the loop linking β6 and β7, a shortening of the loop following β4 and an additional α-helix (helix α1) in the CF I_m25 structure. The additional α-helix in CF I_m25 plays a role in sequestering the

Figure 4. Complex of CF I$_m$25 and diadenosine tetraphosphate. (**A**) The arrow points to the position of Ap$_4$A within the CF I$_m$25 dimer shown rotated 120° as compared to Figure 2. (**B**) The ordered moieties of Ap$_4$A (adenine base and triphosphate) are superimposed on an original, unbiased 1.80 Å (Fo–Fo) isomorphous difference Fourier map contoured at 2.6σ (cyan) and a simulated annealing omit map contoured at 2.7σ (magenta). The adenine base of CF I$_m$25 is stabilized through base stacking interaction with Phe103. Arg63, Arg150, Gln157 and Lys172 contact the oxygens of the β and γ phosphates.

substrate binding pocket from solvent exposure. When comparing the positions of the bound nucleotides between CF I$_m$25 and DR1025 we also see variations in substrate fit. Superposition of the ATP bound in DR1025 on to the CF I$_m$25 structure shows that both the base and phosphate tail would clash with protein residues in helix α1 and beta strand β5 in the CF I$_m$25 structure (Figure 6). Commonly, variations in the nucleotide-binding region of Nudix proteins occur due to differences in the side chains and motifs and contribute to the substrate specificity of the Nudix protein. Interestingly, both DR1025 and CF I$_m$25 stabilize the adenine base via stacking interactions with a phenylalanine residue found outside of the Nudix box region. Base-stacking interactions via a Tyr or Phe located 17 amino acids downstream of the Nudix box is commonly found in all of the Nudix Ap$_4$A hydrolases, including DR1025 (11). Phe103 stabilizes the adenine base in the Ap$_4$A bound CF I$_m$25 structure but is located six residues upstream of the Nudix box region.
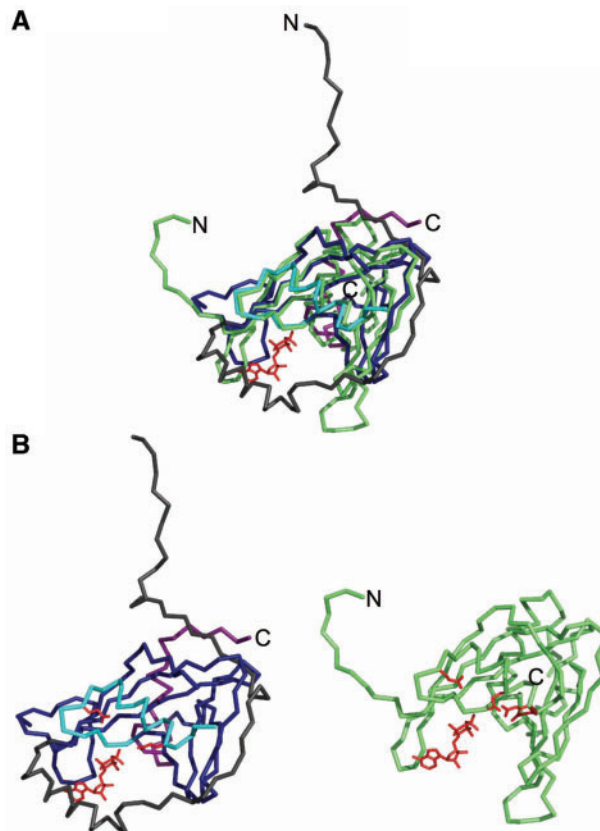
Figure 5. Comparison with other Nudix proteins. (**A**) Superposition of the Nudix regions between DR1025 (PDB ID code 1SU2) (12) and CF I$_m$25 in their monomeric form. The superposition was calculated over residues 12–145 in DR1025 and 77–202 in CF I$_m$25. DR1025 is shown in light green and its bound substrate (ATP) is colored in red. CF I$_m$25 is colored according to its domain architecture as shown in Figure 2. (**B**) Side-by-side representation of the monomeric forms of the substrate bound DR1025 and CF I$_m$25 with ATP from the DR1025 structure and the conserved glutamate residues both shown in red.

### Electrostatic surface and putative RNA and protein-binding regions

The electrostatic surface potential of CF I$_m$25 was calculated with GRASP (Figure 7A) (35). The amino terminal region and residues 78–160 of the Nudix domain were shown to participate in RNA binding (7). We observe a good correlation between the location of the putative RNA binding residues and that of the positively charged residues on the surface of the protein (Figure 7B).

DISCUSSION

In our crystal structure CF I$_m$25 is a homodimer. A dimeric state for CF I$_m$25 is consistent with dynamic light scattering (DLS) and gel filtration experiments performed with the 25 kDa subunit both unliganded and with ATP or Ap$_4$A. The dimeric structure of CF I$_m$25 suggests that the active form of CF I$_m$ may be a heterotrimer composed of a CF I$_m$25 homodimer and either the 59, 68 or 72 kDa subunit. This is consistent with interaction studies in *C. elegans* (DIP interaction database accession DIP:25083N) where the CF I$_m$25 homolog (Uniprot accession Q93716) was

Figure 6. Positions of DR1025 ATP and Ap$_4$A in the CF I$_m$25 structure. Positions of the DR1025 ATP shown in red and CF I$_m$25 Ap$_4$A colored in orange (phosphates) and blue/green (adenine base) within the CF I$_m$25 structure. Ligand position was determined by superposition of the Nudix region over residues 12–145 in DR1025 (PDB ID code 1SU2) and 77–202 in CF I$_m$25. Only one CF I$_m$25 monomer is shown.
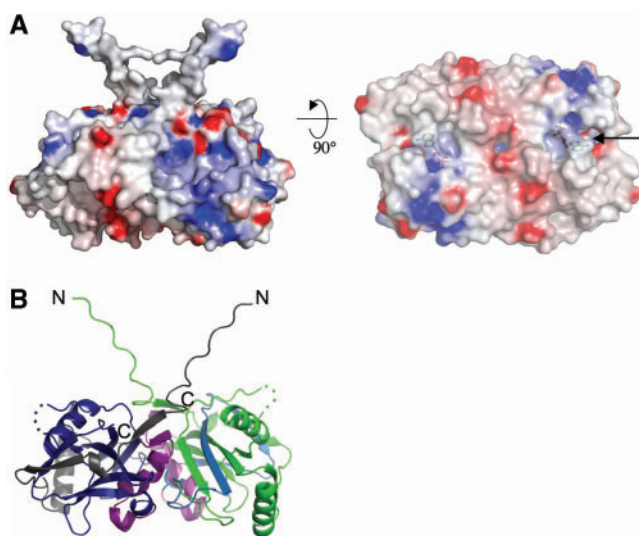
Figure 7. Surface representation of CF I$_m$25. (**A**) Electrostatic potential surface representation of dimeric CF I$_m$25 calculated by the program GRASP (32). The surface is shown in the same orientation as in Figure 2 (left) and rotated 90° (right). The surface on the right was made semi-transparent to show the bound Ap$_4$A (shown with arrow). The areas of negative charge are depicted in red, whereas the positively charged regions are colored in blue. (**B**) Ribbon diagram of dimeric CF I$_m$25 in the same orientation as in **A**, with residues 21–160 corresponding to the RNA-binding region highlighted in green.

found to interact with itself and with the CF I$_m$68 homolog (Uniprot accession Q18937). Alternatively, the complex with the larger subunit of CF I$_m$ could be a heterotetramer. A third possibility is that CF I$_m$25 monomerizes upon binding the larger subunit.

The Nudix box of CF I$_m$25 lacks two of the four glutamates important for catalytic function and metal binding. Our structural and biochemical characterization suggests that CF I$_m$25 is able to bind but not hydrolyze nucleotide substrates. In the dinucleotide bound CF I$_m$25 structure, the dinucleotide is found outside of the Nudix box and lies deeper within the active site compared to ATP in DR1025. The fluorescence-binding data for CF I$_m$25, indicate a binding affinity in the low micromolar range for Ap$_4$A indicative of a potential role as a signaling molecule and a weak, albeit physiologically relevant, binding affinity for ATP. Even though CF I$_m$25 binds both Ap$_4$A and ATP, we have so far only been able to obtain a co-crystal complex with Ap$_4$A. It is entirely possible that the crystallization conditions we have explored to date for the protein–ATP complex are not compatible with the formation of a crystal lattice and that further exploration of the crystallization space might yield the desired conditions. We note that pre-formed crystals of CF I$_m$25 dissolve upon addition of ATP, indicating a possible conformation change upon binding of the nucleotide.

To our knowledge, CF I$_m$25 is the first example of a Nudix protein binding and not hydrolyzing a nucleotide substrate. This loss of function/gain of a regulatory role is not unprecedented in evolution. A similar loss of function was reported for the *Lactobacillus lactis* ATP phosphoribosyl transferase (ATP-PRT) regulatory subunit, His Z (36). ATP-PRT functions to initiate the biosynthesis of histidine and requires both the HisZ subunit and HisG, the catalytic subunit, for activity. The regulatory subunit, HisZ, resembles the catalytic domain of functional histidyl-tRNA synthetases (HisRS) and utilizes its fold, not for catalysis, but for binding of histidine to monitor histidine levels. Although there were no metals evident in either the free or bound structures of CF I$_m$25 and no identifiable hydrolytic activity in our colorimetric assay it remains possible that the Nudix box of CF I$_m$25 could potentiate hydrolysis of Ap$_4$A and ATP upon interaction of CF I$_m$25 with one of the larger CF I$_m$ subunits or additional binding partners.

Ap$_4$A belongs to the family of diadenosine oligophosphates, Ap$_n$A, which were first discovered 40 years ago (37,38). Ap$_4$A is composed of two adenosines and four phosphates linked in 5′ to 5′-phosphodiester linkages. The role of Ap$_n$As in higher eukaryotes has remained elusive. Recently, Ap$_n$As have been suggested to play a role as putative extra- and intracellular signaling molecules. Ap$_4$A itself is involved in the cellular stress response, inhibition of K$_{ATP}$ channels, stimulation of DNA replication and repair, as well as influencing other essential cellular processes in eukaryotes (37,38). Presently, the only link between Ap$_4$A and 3′-end processing was described in yeast where stimulation of primer independent synthesis by yeast poly(A) polymerase was observed in the presence of dinucleoside polyphosphates, including Ap$_4$A (39). There is no CF I$_m$25 homolog in yeast but this observation still suggests a potential role for dinucleoside polyphosphates as signaling molecules during RNA-processing events. The concentration of ATP and Ap$_4$A can fluctuate within the cell in response to cellular stress or growth.

Preliminary experiments on the role of Ap$_4$A during the polyadenylation step of 3′-end processing resulted in a non-competitive inhibition of polyadenylation in a poly (A) extension assay with mammalian PAP (results not shown). This suggests that under conditions of high concentrations of Ap$_4$A such as stress, Ap$_4$A can bind to a site in PAP, other than the ATP binding site, to inhibit polyadenylation.

The residues 81–160 of the Nudix domain play a dual role, binding RNA and stabilizing protein–protein interactions. Additionally, the amino terminus (1–76) participates in RNA binding (Figure 2) (7). A stable interaction between CF I$_m$25 and CF I$_m$68, unlike that involving PAP or PABPN1, requires the entire CF I$_m$25 protein. CF I$_m$68 must contact regions outside of the known RNA-binding region of CF I$_m$25 to promote complex formation and stimulate pre-mRNA 3′-end processing. Interestingly, CF I$_m$25 has a patch of negatively charged surface residues that runs the length of the dimer interface (Figure 7A). This charged region is composed of residues primarily from beta strand β3 and helix α4 and could potentiate protein–protein interactions with CF I$_m$25's other binding partners, possibly CF I$_m$68. Also intriguing is the observation that a binding interaction between the substrate RNA and the 25 kDa subunit occurs in the absence of a putative RNA-binding domain. This suggests another mechanism of RNA recognition, possibly through homodimer formation. A CF I$_m$25 homodimer may enhance the binding potential to the RNA substrate compared to a monomer interaction by increased surface area interactions. A definite answer regarding the oligomeric state of the 25 kDa subunit and the RNA-binding mechanism will have to await a CF I$_m$25 structure with RNA bound.

The results reported here suggest that CF I$_m$25 is unable to hydrolyze nucleotides or dinucleotides even though it harbors a classic Nudix fold. The Nudix domain of CF I$_m$25 may instead facilitate protein–protein interactions, as suggested by the large distribution of charged residues in the electrostatic surface representation of the Nudix domain. This charge distribution of the Nudix domain correlates well with results from pull-down experiments with PAP and PABPN1 (7).

The interaction of CF I$_m$ with the RNA substrate and with PAP stimulates the rate of polyadenylate tail synthesis. This may be facilitated by the binding of ATP to CF I$_m$25. Preliminary data (S. Dettwiler and W. Keller, unpublished) showed that CF I$_m$25 interacts with hClp1, a protein shown to bind ATP which is involved in 3′-pre-mRNA processing and tRNA splicing (40,41). This observation, in conjunction with the fact that hClp1 has recently been shown to function as an siRNA kinase and a kinase that phosphorylates the 5′-end of the 3′-splicing product in human tRNA splicing (42), suggests a possible link between CF I$_m$25's binding of ATP and protein–protein cross talk. Recently, CF I$_m$25 has also been shown to be associated in a large RNP complex with Rae1, an mRNA export protein, in the nucleation and stabilization of microtubules during spindle assembly (43). This interaction is via CF I$_m$25 association with an RNA component of the RNP complex and suggests a role for CF I$_m$25 in mRNA export via direct RNA association that

may be influenced by the concentration of intracellular nucleotides or dinucleotides.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

REFERENCES

1. Zhao,J., Hyman,L. and Moore,C. (1999) Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.*, **63**, 405–445.
2. Wahle,E. and Rüegsegger,U. (1999) 3′ End processing of pre-mRNA in eukaryotes. *FEMS Microbiol. Rev.*, **23**, 277–295.
3. Minvielle-Sebastia,L. and Keller,W. (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr. Opin. Cell Biol.*, **11**, 352–357.
4. Brown,K.M. and Gilmartin,G.M. (2003) A mechanism for the regulation of pre-mRNA 3′ processing by human cleavage factor Im. *Mol. Cell*, **12**, 1467–1476.
5. Venkataraman,K., Brown,K.M. and Gilmartin,G.M. (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev.*, **19**, 1315–1327.
6. Rüegsegger,U., Beyer,K. and Keller,W. (1996) Purification and characterization of human cleavage factor Im involved in the 3′ end processing of messenger RNA precursors. *J. Biol. Chem.*, **271**, 6107–6113.
7. Dettwiler,S., Aringhieri,C., Cardinale,S., Keller,W. and Barabino,S.M. (2004) Distinct sequence motifs within the 68-kDa subunit of cleavage factor Im mediate RNA binding,

protein-protein interactions, and subcellular localization. *J. Biol. Chem.*, **279**, 35788–35797.

8. Millevoi,S., Loulergue,C., Dettwiler,S., Karaa,S.Z., Keller,W., Antoniou,M. and Vagner,S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. *EMBO J.*, **25**, 4854–4864.

9. Bessman,M.J., Frick,D.N. and O'Handley,S.F. (1996) The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.*, **271**, 25059–25062.

10. Mildvan,A.S., Xia,Z., Azurmendi,H.F., Saraswat,V., Legler,P.M., Massiah,M.A., Gabelli,S.B., Bianchet,M.A., Kang,L.W. *et al.* (2005) Structures and mechanisms of Nudix hydrolases. *Arch. Biochem. Biophys.*, **433**, 129–143.

11. McLennan,A.G. (2006) The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.*, **63**, 123–143.

12. Ranatunga,W., Hill,E.E., Mooster,J.L., Holbrook,E.L., Schulze-Gahmen,U., Xu,W., Bessman,M.J., Brenner,S.E. and Holbrook,S.R. (2004) Structural studies of the Nudix hydrolase DR1025 from *Deinococcus radiodurans* and its ligand complexes. *J. Mol. Biol.*, **339**, 103–116.

13. Tropea,J.E., Cherry,S., Nallamsetty,S., Bignon,C. and Waugh,D.S. (2007) In Doublié,S. (ed.), *Macromolecular Crystallography Protocols.* Humana Press, Totowa, Vol. 1, pp. 1–20.

14. Doublié,S. (2007) In Doublié,S. (ed.), *Macromolecular Crystallography Protocols.* Humana Press, Totowa, Vol. 1, pp. 91–108.

15. Otwinowski,Z. and Minor,W. (1997) *Macromolecular Crystallography, Part A.* Academic press, New York, Vol. 276, pp. 307–326.

16. Terwilliger,T.C. and Berendzen,J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **55**, 849–861.

17. Vonrhein,C., Blanc,E., Roversi,P. and Bricogne,G. (2007) In Doublié,S. (ed.), *Macromolecular Crystallography Protocols.* Humana Press, Totowa, Vol. 1, pp. 215–230.

18. Terwilliger,T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.*, **56**, 965–972.

19. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

20. Brunger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M. *et al.* (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.

21. Laskowski,R.A., McArthur,M.W., Moss,D.S. and Thornton,J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, **26**, 283–291.

22. Rould,M.A. (2006) The same but different: isomorphous methods for phasing and high-throughput ligand screening. *Methods Mol. Biol.*, **364**, 159–182.

23. DeLano,W.L., The PyMOL Molecular Graphics System (2002) on the World Wide Web. http://www.pymol.org. San Carlos, CA, USA.

24. CCP4. (1994) The CCP4 suite: programs for protein crystal-lography. *Acta Crystallogr. D Biol. Crystallogr.*, **50**, 760–763.

25. Ames,B.N. and Dubin,D.T. (1960) The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. *J. Biol. Chem.*, **235**, 769–775.

26. Yengo,C.M., Chrin,L.R., Rovner,A.S. and Berger,C.L. (2000) Tryptophan 512 is sensitive to conformational changes in the rigid relay loop of smooth muscle myosin during the MgATPase cycle. *J. Biol. Chem.*, **275**, 25481–25487.

27. Shimazu,T., Horinouchi,S. and Yoshida,M. (2007) Multiple histone deacetylases and the CREB-binding protein regulate pre-mRNA 3′ end processing. *J. Biol. Chem.*, **282**, 4470–4478.

28. Rush,J., Moritz,A., Lee,K.A., Guo,A., Goss,V.L., Spek,E.J., Zhang,H., Zha,X.M., Polakiewicz,R.D. *et al.* (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nat. Biotechnol.*, **23**, 94–101.

29. Holm,L. and Sander,C. (1997) Dali/FSSP classification of three-dimensional protein folds. *Nucleic Acids Res.*, **25**, 231–234.

30. Bailey,S., Sedelnikova,S.E., Blackburn,G.M., Abdelghany,H.M., Baker,P.J., McLennan,A.G. and Rafferty,J.B. (2002) The crystal structure of diadenosine tetraphosphate hydrolase from *Caenorhabditis elegans* in free and binary complex forms. *Structure*, **10**, 589–600.

31. Yoshiba,S., Ooga,T., Nakagawa,N., Shibata,T., Inoue,Y., Yokoyama,S., Kuramitsu,S. and Masui,R. (2004) Structural insights into the *Thermus thermophilus* ADP-ribose pyrophospha-tase mechanism via crystal structures with the bound substrate and metal. *J. Biol. Chem.*, **279**, 37163–37174.

32. She,M., Decker,C.J., Chen,N., Tumati,S., Parker,R. and Song,H. (2006) Crystal structure and functional analysis of Dcp2p from *Schizosaccharomyces pombe*. *Nat. Struct. Mol. Biol.*, **13**, 63–70.

33. Gabelli,S.B., Bianchet,M.A., Bessman,M.J. and Amzel,L.M. (2001) The structure of ADP-ribose pyrophosphatase reveals the structural basis for the versatility of the Nudix family. *Nat. Struct. Biol.*, **8**, 467–472.

34. Logan,D.T., Andersson,J., Sjoberg,B.M. and Nordlund,P. (1999) A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science*, **283**, 1499–1504.

35. Nicholls,A., Sharp,K.A. and Honig,B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins*, **11**, 281–296.

36. Champagne,K.S., Sissler,M., Larrabee,Y., Doublie,S. and Francklyn,C.S. (2005) Activation of the hetero-octameric ATP phosphoribosyl transferase through subunit interface rearrangement by a tRNA synthetase paralog. *J. Biol. Chem.*, **280**, 34096–34104.

37. Kisselev,L.L., Justesen,J., Wolfson,A.D. and Frolova,L.Y. (1998) Diadenosine oligophosphates (Ap(n)A), a novel class of signalling molecules? *FEBS Lett.*, **427**, 157–163.

38. Baxi,M.D. and Vishwanatha,J.K. (1995) Diadenosine polyphosphates: their biological and pharmacological significance. *J. Pharmacol. Toxicol. Methods*, **33**, 121–128.

39. Sillero,M.A., De Diego,A., Osorio,H. and Sillero,A. (2002) Dinucleoside polyphosphates stimulate the primer independent synthesis of poly(A) catalyzed by yeast poly(A) polymerase. *Eur. J. Biochem.*, **269**, 5323–5329.

40. Noble,C.G., Beuth,B. and Taylor,I.A. (2007) Structure of a nucleotide-bound Clp1-Pcf11 polyadenylation factor. *Nucleic Acids Res.*, **35**, 87–99.

41. de Vries,H., Rüegsegger,U., Hübner,W., Friedlein,A., Langen,H. and Keller,W. (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J.*, **19**, 5895–5904.

42. Weitzer,S. and Martinez,J. (2007) The human RNA kinase hClp1 is active on 3′ transfer RNA exons and short interfering RNAs. *Nature*, **447**, 222–226.

43. Blower,M.D., Nachury,M., Heald,R. and Weis,K. (2005) A Rae1-containing ribonucleoprotein complex is required for mitotic spindle assembly. *Cell*, **121**, 223–234.

44. Pearson,W.R. and Lipman,D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. U.S.A*, **85**, 2444–2448.

45. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.